

Sablayrolles Jean-François (Paris 13, LDI UMR 7187) et Cartier Emmanuel (Paris 13, LDI UMR 7187)

« Nouvelles technologies, nouveaux modèles linguistiques et néologie »

Sciences du langage et nouvelles technologies. Actes du colloque 2009 de l'Association des Sciences du Langage, Ponchon Thierry et Laborde-Milaa Isabelle (éd), Limoges, Éditions Lambert-Lucas, 2011, p. 53-59.

Résumé

L'équipe néologie du laboratoire LDI a eu besoin d'une base de données non seulement pour engranger ses « prises » dans sa tâche de veille néologique, mais aussi pour analyser les néologismes relevés et l'utiliser comme moyen d'étude de la néologie du français contemporain. La base Neologia a d'emblée été conçue comme un véritable outil de travail. Une collaboration étroite et régulière entre les lexicologues et le linguiste-informaticien a permis de construire cette base qui répond aux objectifs fixés. Après une période de rodage d'un peu plus d'un an, la base a été corrigée en fonction des lacunes que son utilisation a permis de mettre au jour. Ses principes et son utilisation seront brièvement présentés.

Le pouvoir heuristique de la base Neologia sur notre travail n'a pas été négligeable. Les discussions lors de la création de la base ont conduit à approfondir la réflexion théorique dans le domaine en nous obligeant à expliciter jusque dans le détail nos objectifs, et son utilisation a permis de découvrir des problèmes auxquels on n'avait pas songé initialement. Ce qui a également contribué à des approfondissements théoriques ou à des précisions rassemblés dans un recueil à usage interne de l'équipe, mais qui ont fait et feront l'objet de communications et articles.

Mots clés : néologie, base de données, classes d'objets

Nouvelles technologies, nouveaux modèles linguistiques et néologie

Introduction

Le lexique et les dictionnaires occupent une place centrale en TAL. Parmi les modèles descriptifs récents en ce domaine, celui des classes d'objets développé par Gaston Gross au LDI -avec la confection de dictionnaires de prédicats et d'arguments explicitant la combinatoire et l'actualisation possibles de tous les éléments de chaque classe- a des incidences sur la conception de la néologie et l'extraction des néologismes. Par ailleurs, les bases de données informatisées permettent non seulement d'engranger des informations, comme les fiches cartonnées d'antan, mais elles sont aussi de véritables outils de recherche qui, par des requêtes simples ou croisées, permettent une exploitation directe et efficace des données pour une meilleure connaissance du phénomène, et assurent aussi le suivi de la diffusion des néologismes.

1. Incidences sur la conception de la néologie

Contrairement à la doxa, ce n'est pas le concept de néologie qui est flou, c'est l'absence de descriptions systématiques de la langue qui pose problème dans le repérage des néologismes : nous ne disposons pas de pierres de touche fiables pour décider de l'existence d'une innovation lexicale, qu'il s'agisse de l'apparition d'une nouvelle lexie¹, inexistante auparavant ou d'un emploi innovant par

¹ Sous ce terme, créé par Pottier et repris par d'autres linguistes dont Jean Tournier comme dénomination de l'unité lexicale, nous entendons un signe linguistique ayant une unité fonctionnelle et mémorisée (ou mémorisable) en compétence.

rapport à la combinatoire attestée d'une unité déjà existante dans la langue. Le recours au(x) dictionnaire(s) traditionnel(s) comme corpus d'exclusion ne peut être mécanique du fait de leurs différences, des lacunes de leurs nomenclatures, de l'insuffisance de leurs descriptions et de l'existence, généralement oubliée, des néologismes homonymiques². Bien sûr ces dictionnaires sont des outils qui apportent nombre d'informations intéressantes, mais leur manque de systématisme les rend difficilement utilisables. C'est là où réside la supériorité de dictionnaires directement informatisés et comportant des informations sur les divers emplois d'un signifiant³. Une lexie trouvée dans un texte et absente comme entrée ou avec un emploi non décrit relève de la néologie, sauf en cas d'une éventuelle lacune du dictionnaire, qu'il faut alors modifier en conséquence. Outre les emprunts et la néologie formelle (morpho-sémantique ou purement morphologique de J. Tournier⁴), il y a la néologie syntactico-sémantique⁵ et la néologie syntaxique (des changements de construction sans modification sensible du sens⁶). La néologie par détournement pose des problèmes spécifiques qui ne seront pas abordés ici.

2. Incidences pratiques 1 : L'extraction automatique des néologismes

L'extraction (semi)automatique des néologismes se fait à l'aide de corpus d'exclusion. Le problème principal repose donc sur la qualité et la fiabilité de ce corpus d'exclusion. S'il se contente de reprendre la nomenclature des dictionnaires papier, on a les mêmes problèmes que ceux déjà évoqués ci-dessus et ailleurs. Il faut donc se confectionner un dictionnaire informatique plus fiable ou en avoir un à sa disposition. C'est le cas au LDI avec le dictionnaire Morfetik élaboré depuis des années par Michel Mathieu-Colas. Il comporte de l'ordre d'un million de formes et le dictionnaire des mots composés, encore indépendant, va y être incessamment intégré. À l'aide de l'outil Telanaute mis au point par Fabrice Issac, nous importons des textes de la presse quotidienne en ligne dont les « mots » sont confrontés au dictionnaire Morfetik. Tout ce qui n'est pas reconnu est présenté comme

² Sur les rapports entre les néologismes et les dictionnaires, voir, entre autres, Sablayrolles « Fondements théoriques des difficultés pratiques du traitement des néologismes », *Revue française de linguistique appliquée*, vol. VII-1. / juin 2002 « Lexique : recherches actuelles », pp. 97-111 et « Néologie et dictionnaire(s) comme corpus d'exclusion », *Néologie et terminologie dans les dictionnaires*, J.-F. Sablayrolles éd., coll. Lexica, H. Champion, 2008, pp. 19-36.

³ On évite volontairement d'employer ici le terme lexie, dans la mesure où cela entraînerait une longue discussion sur le traitement polysémique ou homonymique des lexies.

⁴ Outre les procédés traditionnels et bien connus de la néologie morpho-sémantique que sont la dérivation et la composition sous leurs diverses formes, il y a, moins connues, les innovations flexionnelles sans changement de sens (*je closis*, *la représsaille*) ou avec des inflexions de sens plus notables (*les banlieues* ou *les quartiers* qui ne s'appliquent qu'à des lieux défavorisés ou à problèmes). La néologie purement morphologique correspond aux diverses troncations et siglaisons.

⁵ Entrent dans cette grande catégorie tous les changements d'emploi qui ne s'accompagnent pas d'une modification affixale : les conversions et la néologie sémantique. Dans les deux cas, il y a infraction par rapport à la combinatoire décrite systématiquement dans des dictionnaires informatiques, ainsi de l'emploi maintenant répandu de l'adjectif *grave* comme adverbe (*ça m'ennuie grave*) ou de tous les emplois figurés comme *formater* pour un humain (*Jean a été formaté pour ce poste*).

⁶ Ce concept et cette dénomination sont anciens et se trouvent exposés, entre autres, dans le *Dictionnaire Universel* de Pierre Larousse et au détour de certaines pages de l'*Histoire de la Langue française* de Ferdinand Brunot. On peut en distinguer au moins deux grandes catégories. La première consiste essentiellement dans des changements dans l'emploi de préposition (suppression : *il craint*, en français non standard, ajout : *pallier à*, remplacement par une autre : *semblable avec*). La seconde se manifeste dans l'apparition d'un prédicat sous une forme catégorielle inattendue : si la forme verbale *prendre* et la forme nominale *prise* sont possibles quand il s'agit des classes d'objets <médicament> ou <lieu institutionnel> (*prendre un médicament* / *la prise de ce médicament* ; *prendre Troie*, *la Bastille* / *la prise de Troie*, *de la Bastille*), seule la forme verbale est normalement attestée pour les <moyens de transports> : *prendre le bus* / **la prise du bus*. C'est donc avec une innovation (par transgression volontaire destinée à frapper le public cible) que des publicitaires ont créé le slogan au bénéfice des TER de la SNCF *La prise de train bénéficie à la santé de votre voiture*.

candidat néologisme. Trois situations se présentent alors. Soit il s'agit de vrais néologismes, et il sont intégrés dans la base Neologia (voir infra), soit ce sont des lacunes de Morfetik et ces lexies sont intégrées dans le dictionnaire, soit, et c'est quantitativement encore le plus fréquent, il s'agit de fautes diverses que l'on élimine d'un coup de souris, mais qui sont intégrées dans le corpus d'exclusion afin de ne pas les voir réapparaître. Petit à petit, le nombre de ces scories tendra à diminuer sans pouvoir être totalement éliminées néanmoins.

Si le système d'extraction des néologismes formels est au point et fonctionne assez bien, l'extraction des autres néologismes, par changement de combinatoire (syntactico-sémantique ou purement syntaxique voir supra), en est encore à ses balbutiements. Les processus de comparaison de la combinatoire effectivement employée dans les textes avec celle, conventionnelle, telle qu'elle devrait être consignée dans les dictionnaires de langue et telle que s'efforcent de la décrire les dictionnaires informatiques élaborés au LDI, sont envisageables dans l'avenir mais beaucoup plus complexes à mettre en œuvre, ne serait-ce que parce que les dictionnaires informatiques sont des dictionnaires de phrases minimales⁷ et que les textes se présentent avec des phrases de structures plus complexes comprenant des prédicats de second degré, des expansions, des insertions, des problèmes d'ordre des constituants, etc. Une phase de transition consiste donc sérier les problèmes et à opérer sur des phrases réduites aux constituants obligatoires.

3. Incidences pratiques 2 : une base de traitement et de suivi des néologismes, Neologia

3.1. Un outil de travail

Engranger « prises » de la veille néologique et aussi analyser les néologismes relevés. La base Neologia a d'emblée été conçue comme un véritable outil de gestion et d'étude des néologismes du français contemporain. Elle comporte une trentaine de champs répartis en cinq groupes : entrée et définition, morphosyntaxe, syntactico-sémantique, néologie, relations sémantiques et contextes.

Les fonctionnalités d'édition de recherche et de consultation implantées dans l'outil permettant des requêtes simples et requêtes multicritères de divers types.

Pouvoir heuristique de la base Neologia sur le travail n'a pas été négligeable. Les discussions lors de la création de la base ont conduit à approfondir la réflexion théorique dans le domaine en nous obligeant à expliciter jusque dans le détail les objectifs, et son utilisation a permis de découvrir des problèmes auxquels on n'avait pas songé initialement. Ce qui a également contribué à des approfondissements théoriques ou à des précisions rassemblés dans un recueil à usage interne de l'équipe, mais qui ont fait et feront l'objet de communications et articles.

3.2. Suivi de la circulation des néologismes jusqu'à la perte de leur statut de néologisme

Informatique fournit un outil pour gérer le cycle de vie de ces mots néologiques : émergence, développement ou disparition, intégration ou non dans les dictionnaires généraux

Outils de suivi de l'utilisation des mots par la communauté linguistique dans les corpus textuels numériques, permettant de gérer le cycle de vie des néologismes : enregistrement des premières occurrences, puis suivi, via différents outils statistiques sur le web, de l'utilisation des mots. De la sorte, nous appliquons un modèle permettant de gérer le dynamisme de la nomenclature linguistique.

Module de gestion des néologismes, qui permet de suivre, une fois l'émergence du néologisme enregistrée, de suivre son évolution au travers des flux textuels accessibles sur internet. Ce suivi permet ensuite de profiler chaque néologisme et de définir un seuil au-dessus duquel un néologisme doit intégrer le dictionnaire général ou spécifique d'une langue.

Conclusion

Collaboration linguistes / informaticiens nécessaire et fructueuse
Beaucoup de projets mais pas assez de temps pour tout faire : des choix et des priorités à établir

⁷ Une phrase minimale est un prédicat saturé par ses arguments.

